

CORRELACIONES

Cuando la Estadística se apodera de un tema es capaz de hincharlo hasta el infinito y su recurso más socorrido para ello es el coeficiente de correlación. Vamos a presentar un ejemplo partiendo de series numéricas sacadas del Calendario Meteoro-Fenológico, pues cualesquiera de ellas sirven, puesto que el método es independiente de la significación de las series utilizadas.

Empecemos por recordar que el coeficiente de correlación es un índice numérico que se calcula apareando dos series del mismo número de términos, reducidas previamente a series de desviaciones respecto de su media aritmética; multiplicando cada dos términos correspondientes y dividiendo el resultado por el producto de las desviaciones tipo. El resultado es siempre menor que la unidad en valor absoluto y el signo puede ser positivo o negativo. Este índice mide el grado de dependencia que liga las dos magnitudes que se comparan: cuanto más se acerca a la unidad más fuerte es el enlace, que se dice directo cuando la correlación es positiva; e inverso cuando es negativa. Cuando el coeficiente de correlación es cero, significa que las citadas magnitudes son independientes: cuando vale la unidad es que dicha relación es lineal. Hemos escogido como

ejemplo los volúmenes de precipitación mensual del año 1974, correspondientes a las cuencas del Duero, Ebro y Pirineo oriental. Los datos iniciales reducidos a desviaciones respecto a la media mensual del mismo año son los siguientes (en millones de metros cúbicos).

	Duero	Ebro	P. Oriental
Enero	+ 3.439	— 1.934	— 614
Febrero	+ 664	— 308	— 395
Marzo	+ 763	+ 6.232	+ 1.197
Abril	— 147	+ 377	+ 512
Mayo	— 535	— 783	+ 247
Junio	+ 3.092	— 296	+ 92
Julio	— 1.765	— 917	— 343
Agosto	— 2.505	+ 130	+ 481
Septiembre	— 2.615	+ 1.360	+ 824
Octubre	— 1.909	— 86	— 294
Noviembre	+ 2.965	— 164	— 696
Diciembre	+ 1.618	— 3.705	— 912
Desviación tipo	1.885	2.226	627

Covarianza Duero-Ebro = 971233.

Coefficiente correlación Duero-Ebro = — 0,231.

Covarianza Duero-P. Oriental = 547641.

Coefficiente correlación Duero-P. Oriental = — 0,463.

Covarianza Ebro-P. Oriental = + 1057862.

Coefficiente correlación Ebro-P. Oriental = + 0,758.

Salta a la vista que el coeficiente de correlación Duero-Ebro no es significativo; en cambio el del Ebro-Pirineo Oriental es muy elevado y además positivo; el de Duero-Pirineo Oriental lo es poco y negativo. La explicación no es difícil. No es que el agua caída en

una cuenca pueda ser causa de la que caiga en otra cuenca vecina, pero ambas pueden ser efecto de una causa común. Las condiciones topográficas y meteorológicas son las que más influyen en la distribución de la lluvia y estas causas si no idénticas pueden ser bastante semejantes en cuencas próximas, como lo son las del Ebro y del Pirineo Oriental. En cambio esta última cuenca y la del Duero, contiguas, están bajo condiciones geográficas distintas, pues una es de la vertiente atlántica y la otra del Mediterráneo y por eso carecen prácticamente de correlación y ésta es de signo negativo. Más difícil de explicar porque el Duero y el Pirineo Oriental, todavía más separadas y de distinta cuenca, muestran un coeficiente de correlación medianamente significativo. Parece ser que en este caso las razones hay que buscarlas en la Meteorología dinámica más que en la Geografía. Las borrascas que penetran en Europa son mucho más extensas que nuestras cuencas hidrográficas y como todas tienen una estructura más o menos similar y se deslizan sobre una latitud que varía poco, podría explicarse el fenómeno suponiendo que la mitad de las depresiones cruzan el norte de la península y la otra mitad el centro: en el primer caso riegan el norte y dejan seco el Duero y en el segundo al revés.

Vamos a ofrecer un ejemplo de otro tipo de correlación, de particular importancia en Climatología: la correlación secuencial. Esta técnica se aplica a una sola serie, apareando cada desviación con la que le sigue y calculando el coeficiente de correlación como si fuesen series paralelas. La desviación tipo en este

caso es naturalmente la misma para ambas series y el denominador será el cuadrado de esta desviación. Escojamos la serie Duero y obtenemos:

Covarianza secuencial Duero = + 1300703.

Correlación secuencial Duero = 0,514.

El resultado tampoco es sorprendente y significa que la serie de precipitaciones de un mismo año en meses sucesivos está poco organizado, pero tampoco puede decirse que sea completamente al azar (el coeficiente de correlación sería casi cero, en este caso). Si en lugar de relacionar la precipitación mensual con la del mes siguiente buscásemos la de un día con la del día siguiente, el resultado sería otro, pues en este caso sin duda existe cierto enlace físico entre ellas. Efectivamente, efectuados los cálculos resulta para el coeficiente de correlación secuencial 0,654. En cambio comparando la precipitación de cada día no con la del día siguiente, sino con la de quince días después obtenemos 0,253, que confirma esta independencia práctica entre estos fenómenos.

Como se ve por los anteriores ejemplos, que podríamos prolongar sin parar, el método del coeficiente de correlación permite descubrir conexiones entre fenómenos meteorológicos, que tal vez quedarían ocultas, o al contrario desacreditar relaciones supuestas que no tienen fundamento objetivo. Este índice por ser un número abstracto, es decir que no depende de las unidades de medida empleadas y por consiguientes tanto

puede aplicarse a series homogéneas como heterogéneas, sin ninguna preparación previa. Incluso puede usarse para puntualizar las influencias que tiene la Climatología con otras disciplinas, como la Agricultura (series de cosechas, etc.), Hidrología (caudales), Sanidad (epidemias y endemias), y no digamos de la utilidad que muestra dentro de la propia Meteorología, descubriendo trabazones internas entre unas magnitudes y otras y comparando índices tan complicados como se quiera.

Apéndice

Cálculo del coeficiente de correlación. Se disponen las dos series a comparar en dos columnas paralelas. Se calcula la media aritmética de cada una sumando sus términos y dividiendo por el número de ellas. El resultado se resta de cada una de ellas, poniendo de manifiesto el signo y las diferencias obtenidas se vuelven a disponer en dos columnas paralelas. Se eleva al cuadrado cada una de estas diferencias y se suman los resultados y se divide por el número de términos (separadamente para cada columna); así se tienen las *varianzas*; su raíz cuadrada es la *desviación tipo*. Por otra parte se multiplica cada diferencia por su correspondiente de la columna y se suman estos productos; dividiendo también por el número de sumandos re-

sulta la llamada *covarianza*. El cociente de la covarianza por el producto de las desviaciones tipo es el *coeficiente de correlación*.

Fórmulas:

$$\text{Varianza} = V(x) = \frac{1}{N} \sum (x - \bar{x})^2$$

$$\text{Desviación tipo} = \delta_x = \sqrt{V}$$

$$\text{Covarianza} = \text{cov}(x, y) = \frac{1}{N} \sum [(x - \bar{x})(y - \bar{y})]$$

$$\text{Coeficiente de correlación} = r(x, y) = \frac{\text{cov}(x, y)}{\delta_x \cdot \delta_y}$$

(x = término de una serie; \bar{x} = media aritmética; y = término de la otra serie; \bar{y} = media aritmética; N = número de términos de cada una de las series).

J. M. JANSÁ

Meteorólogo